Naval Research Laboratory

Washington, DC 20375-5320

NRL/MR/5540--10-9292

Incentivized Cloud Computing: A Principal Agent Solution to the Cloud Computing Dilemma

Anya Kim
Ira S. Moskowitz

Center for High Assurance Computing Systems
Information Technology Division

September 15, 2010

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
15-09-2010	Memorandum Report	1 April - 25 April 2010

4. TITLE AND SUBTITLE

Incentivized Cloud Computing: A Principal Agent Solution to the Cloud Computing Dilemma

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
61153N

6. AUTHOR(S)

Anya Kim and Ira S. Moskowitz

5d. PROJECT NUMBER

5e. TASK NUMBER
IT015-09-41

5f. WORK UNIT NUMBER
4288

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Naval Research Laboratory, Code 5540
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER

NRL/MR/5540--10-9292

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
One Liberty Center
875 North Randolph Street, Suite 1425
Arlington, VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)
ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

We propose the paradigm of Incentivized Cloud Computing to examine ways of addressing the dilemma of a cloud consumer wanting to utilize cloud resources, but at the same time having to risk the security of its data and applications by trusting a cloud provider to properly manage them. In particular, we look at the theory of incentives, from the areas of economics, health care and public policy, as a means to strengthen the trust relationship between a cloud consumer and cloud provider.

15. SUBJECT TERMS

Cloud computing
Security
Principal agent theory
Moral hazard
Trust and risk

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Anya Kim
Unclassified	Unclassified	Unclassified	UL	16	19b. TELEPHONE NUMBER *(include area code)* (202) 767-6698

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

Incentivized Cloud Computing: A Principal Agent Solution to the Cloud Computing Dilemma

Anya Kim & Ira S. Moskowitz
Center for High Assurance Computing Systems
Information Technology Division
Naval Research Laboratory
Washington, DC 20375

Abstract

We propose the paradigm of *Incentivized Cloud Computing* to examine ways of addressing the dilemma of a cloud consumer wanting to utilize cloud resources, but at the same time having to risk the security of its data and applications by trusting a cloud provider to properly manage them. In particular, we look at the theory of incentives, from the areas of economics, health care and public policy, as a means to strengthen the trust relationship between a cloud consumer and cloud provider. We present this theory, known as principal agent theory, as a new paradigm, in contrast to the existing means of technical and policy-based solutions. Our paradigm examines ways to increase trust in the cloud provider by incentivizing it to share some risk with the consumer. We take a standard principal agent model and discuss the various factors involved in the model, as well as ways to apply it to a consumer-provider relationship in cloud computing.

1 Introduction

Cloud computing is being touted as the next big thing in the IT industry, purportedly offering unmatched levels of scalability, flexibility, just-in-time delivery, and ease of management burdens [26]. Some even claim that cloud computing itself is a paradigm shift, following the shift from the mainframe to client-server model [34]. However, security concerns and loss of control issues associated with cloud computing make it difficult to trust a cloud provider[1]. While cloud consumers are freed from management burdens, they now must trust the provider to perform these management duties satisfactorily. Cloud providers, as revenue generating entities, are motivated to increase their profit, which does not necessarily imply good customer service. There have been several incidents already, where due to poor management from the provider, valuable user data was lost [10], and access to consumer data in the clouds was unavailable [27] when needed. Furthermore, once a contractual relationship has been formed between the consumer and provider, it is non-trivial for the consumer to migrate its applications and data to another provider, if the current level of service proves unsatisfactory. Therefore, the need to choose the correct provider (i.e., one that will provide the desired level of service) is a concern if one wishes to consume cloud computing services.

While various technical and policy-based mechanisms are being examined to enhance the trust in a provider (contracts, monitoring tools, etc.), we argue that these mechanisms are not sufficient by themselves to ensure that the provider will perform according to the wishes of the consumer. While the current paradigm for information security is to secure data and infrastructure using all possible technical means and standardized methodologies, our argument is that "when the cat's away, the mice will play" ... no matter how many "official standard" mouse traps are set. Instead of mouse traps, we need to bribe the mice with

[1]For now, let us just define trust as the belief in another party's intentions. If it does not cause too much agitation, we would also like to state that trust and risk are strongly related concepts.

Manuscript approved August 3, 2010.

a sufficient amount of cheese — enough that they will accept the task, but not so much that they will get lazy. That, in a nutshell is our view of Incentivized Cloud Computing.

The use of economic models to address various computer security issues is not novel. Blakley [9] presents a privacy protection system that does not depend on secrecy, but rather an economic model of self-interest satisfying behavior and points to Axelrod's work in using incentives for cooperation [6]. In his discussion of trust, Jøsang [20] defines trust as knowledge about security. He goes on to say that "gathering as much knowledge as one can about the system, a user will get an idea about the security, or in other words, a certain trust in the system." We can take this one step further and point out that the information asymmetry due to the user's lack of knowledge creates the trust problem, and that with full knowledge of the security of a system, there would be no hidden information and thus perfect trust in the system.

We propose a way to "incentivize" the provider to make security management choices that are beneficial to the consumer. In other words, we need to find a way to create incentives that will compel the provider to take actions that align with the (security) objectives of the consumer. Our approach uses models and methods developed in economics (in particular, landlord-tenant or employer-employee relationships) to act as guidelines for cloud computing relationships as a way to supplement or enhance current contracts between cloud consumers and providers. Specifically, principal agent theory from the fields of economics, public policy, insurance, and political science, is well-suited for analyzing the issues associated with the delegation of authority and the resulting loss of control [31, 15, 14]. In particular, the problem of motivating a party to act on behalf of another is known as the principal agent problem. The problem arises when there is information asymmetry, and/or risk-averse behavior (i.e., conflicting goals), and/or uncertainty about the outcome.

This paper is organized as follows. Section 2 offers some background on cloud computing and principal agent theory. Cloud computing is covered very briefly, and principal agent theory is expounded upon. Section 3 sets out our paradigm in detail by describing how the principal agent problem can be applied to cloud computing by creating an incentive structure that can be applied to cloud computing. At the end of section 3, we present the our paradigm within a game theory framework. Section 4 is a section designed to invite discussion and feedback. In section 5, we conclude by analyzing various issues and limitations that must be considered.

2 Background

2.1 Cloud Computing

Cloud computing (see Fig. 1) is a relatively new buzz phrase for a relatively old technology that has almost as many definitions as there are papers about it [33]. For the purposes of this paper, we do not attempt to redefine or add another definition to the vast pool of definitions but rather lamely quote Wikipedia in which cloud computing is defined as [34], "a model for enabling available, convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction." In our opinion, the above definition covers the salient features of cloud computing that are addressed in this paper. In cloud computing, the resources available are scalable and highly virtualized, enabling users to utilize and pay for only as much computing power as they need.

Many entities are involved in a cloud computing environment. We are interested specifically in the cloud provider and the cloud consumer. The cloud provider is the entity which owns and manages the resources. The cloud consumer is the entity that consumes the resources and may be an individual or an organization (as depicted in Fig.1). When a cloud consumer is an organization, it will have users or employees that access cloud resources. There may also be casual users that have no relationship with the cloud provider, but are accessing cloud consumers' web services that are developed and hosted within the cloud. The trust relationship of concern here is between the two entities that are part of the contract: in other words, the cloud provider and the cloud consumer.

As discussed, there are several papers and magazine articles that discuss cloud computing in relative

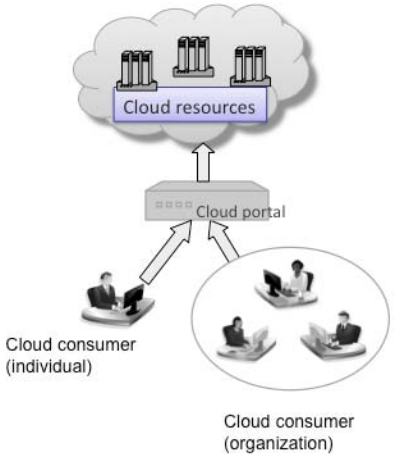

Figure 1: Simplified Depiction of Cloud Computing

detail. Interested readers should refer to [4, 25, 22], while we provide only a simple background that is relevant to this paper. The major touted benefits, or characteristics, of cloud services are: on-demand self-service, ubiquitous network access, location independent resource pooling, rapid elasticity, and measured service [26]. There are three basic delivery models for cloud computing as defined by NIST [26]:

- Infrastructure-as-a-Service (IaaS): provides the infrastructure (computing platform), resources and tools
  (servers, storage, network, etc) to build an application environment. Amazon's EC2 is an example of an IaaS.
- Platform-as-a-Service (PaaS): provides the computing platform as well as solution stack for consumers to develop their own applications and host their own data. Google Apps is one of the major PaaS providers.
- Software-as-a-Service (SaaS): provides the computing platform and applications to customers for use. Common examples are Facebook, Twitter, and various web-based email systems such as those offered by Google.

In general, the less the cloud provider offers in its delivery model, the more flexibility the consumer has, as well as more responsibility in implementing and maintaining security within the cloud [13]. For instance, in an IaaS deployment, the provider only secures the infrastructure and platforms; the security of the higher level components such as virtual instances, operating systems and user data are the responsibility of the consumer. However, regardless of the deployment model, the cloud provider's management of the underlying infrastructure affects the health of the layers above it, thus affecting the security of consumer-owned or consumer-managed components. Furthermore, the provider must create and maintain common appropriate security measures such as authentication and access control to the cloud portal, load balancing, and backup regardless of the model used. When these mechanisms fail, the security of the consumer's data is at risk. The common concern raised in adopting a cloud computing framework is the need for (or lack of) security and the amount of trust required [11, 12, 16, 18, 19]. In fact, the security issues that exist in traditional systems still exist here, but additionally, new problems associated with the multi-tenancy characteristics of cloud computing are present as well [7].

Cloud computing can be further categorized depending on the ownership of the infrastructure: Private clouds are owned and maintained by the same entity and public clouds are offered to the general public while managed by a cloud provider. In a private cloud owned and managed by the same entity, there is no trust issue to consider, hence loss of control and associated risk are not issues. It will come as no surprise then, that the majority of identified security concerns are not present in private clouds. The concept of virtual private cloud (VPC) has been proposed by cloud providers as a way to offer the security of private clouds while still providing the scalability and flexibility offered in public clouds. The terminology is deceiving - VPCs are still public clouds, with dedicated lines connecting consumers to "isolated" resources within the cloud [2]. As long as cloud management is performed by a party other than the consumer, risks emerge and a trust relationship is required. Hence, public clouds and VPCs both are of concern to us from the risks associated with third party management.

2.2 Principal Agent Theory

When one party - the principal - delegates a task to another party - the agent, a principal agent relationship is established. In cloud computing, a principal agent relationship exists between the cloud consumer (principal) and the cloud provider (agent). Incentive problems may arise in these relationships. The problem between the two parties arises due to three factors: conflicting objectives, information asymmetry, and difficulty of monitoring the agent [8]. The presence of any one of these conditions can trigger the incentive problem. Therefore, the ideal way to solve the problem is to eliminate or minimize the effects of these factors:

- Align the objectives: as long as cloud providers are companies whose sole aim is to make a profit, or increase their net value, we cannot truly expect them to share a common goal with their (many) cloud consumers. In the real world, agents would prefer to work less, while principals desire the agent's best effort. If in a futuristic world, we can create automatons that serve as cloud providers, we can program them to align their goals to ours, and then the principal agent problem will disappear for good. Until that time, we must be resigned to the fact that differing goals is an inherent problem.
- Eliminate information asymmetry: a key assumption in the literature is that the agent has comparatively more information and expertise about the task than the principal does. This makes it difficult not only to select the appropriate agent for the job, but also allows agents to capitalize on the information advantage by pursuing their own self-interests at the expense of the principal. While the cloud computing literature claims that one of the benefits of using a cloud provider is that they have more knowledge about the system than many organizations, the principal agent literature would pose it as a severe disadvantage to the cloud consumer. Depending on the type of consumer, the degree of information asymmetry will be different, but since only the provider is fully aware of the architecture of its system, there will always be some degree of information asymmetry.
- Monitor the agent's actions: Given the possibility that the agent may cheat or shirk responsibility, the principal requires accountability. Monitoring the agent can be costly and in a cloud computing setting, takes away from the advantage of third party management that is a touted benefit of cloud computing. Besides, perfect monitoring of the agent is impossible. Instead, principals must use imperfect estimators of the agent's effort, such as monitoring of performance or output.

It can be seen that in cases of delegation, solving the incentive problem is not trivial. The problem may be alleviated if some monitoring is possible, some information asymmetry can be lessened and/or the principal can entice the agent to at least partially share its objectives. Principal agent theory arose in the field of organizational economics [31] to examine ways of resolving problems by enforcing certain conditions through a contract or incentive scheme. The incentive scheme attempts to have the agent accept some risk, so that it will be motivated to work harder for the principal. While this does not totally align the goals of the two parties, it compels the agent to act as if they do. The theory examines various problems such as selecting the right third party for the task, the difficulty in motivating the third party to work on behalf of the user, and the lack of the agent's willingness to perform well, among others. Since perfect monitoring is not feasible these incentive schemes are based on the verifiable outcome.

The two main problems associated with principal agent theory and information asymmetry in particular are moral hazard (hidden action) and adverse selection (hidden information) [21, 5]. Adverse selection arises prior to accepting the contract and is concerned with the difficulty of selecting the appropriate agent from a pool of many potential agents. Moral hazard problems arise after the contract has been accepted, and looks at ways to motivate the agent to accept a degree of risk and perform at an effort level desired by the principal. While both problems are of interest and applicable to a cloud computing consumer-provider scenario, in this paper we focus on the moral hazard problem, i.e., how to get the cloud provider to exert a certain level of effort.

In the case of moral hazard, the principal's goal is to devise an incentive scheme which the agent will not shirk. At a minimum, the incentive scheme should satisfy two conditions. First, the agent must be willing to accept the incentive scheme. In other words, the incentive scheme must offer the agent at least as much utility (i.e., satisfaction) as the agent's next best alternative, also known as its threshold wage or reservation utility. For example, this reservation utility is the compensation that the agent could receive by performing some other task, or working for someone else rather than work for the principal. Second, the scheme should induce the agent to provide the level of effort that the principal desires. In other words, upon accepting the incentive scheme, the agent must be willing to comply with its terms. It is obvious from these conditions that we make assumptions that both the principal and agent adopt maximizing behavior and maximize their individual utility (i.e., from an economics perspective, both are rational).

An incentive scheme that compels the agent to take the most efficient action and accept the threshold wage is called Pareto-optimal [8] (or first-best efficient) for the principal. However, because it is assumed that agents are risk-averse, contracts that offer a threshold wage are generally not sufficient. The principal needs to provide a risk premium above and beyond the expected payment. The magnitude of the risk premium depends on the agent's degree of risk aversion. This type of incentive scheme is called a second-best incentive scheme. It attempts to maximize the principal's utility subject to the following constraints.

- The incentive scheme offers the agent a utility that is at least as high as its threshold wage
- The incentive scheme compels the agent to provide the level of effort most desired by the principal

Other factors that may affect the incentive structure include the effort aversion of the agent and the marginal contribution of effort to profitability [8]. Needless to say, the principal agent problem can be stated mathematically. In the principal agent literature, many mathematical models exist. Some are discrete, others continuous. Some are extremely simplified, some are overly complex. Others assume risk neutrality of both parties. Then others assume various different risk attitudes for the principal and agent. For our purposes, we choose a standard principal agent model as provided by Holmström [17]. When the agent performs a task on behalf of the principal, it is assumed that the agent will exert a level of effort a that will produce output x. This output depends on the agent's level of effort as well as random factors outside its control. In other words, the agent's non-negative output (outcome or payoff) x is assumed to be a continuous random variable with distribution $F(x, a)$, and probability density function $f(x, a)$. Since we assume that the principal cannot monitor the effort, but only the output, the incentive scheme needs to be proportional to the agent's output. At the same time, the principal will try to maximize their own expected utility (i.e., pay as little as possible) that they receive from the agent's output, which is the total output x less the incentive scheme given to the agent. Lastly, for the agent to accept this scheme, the principal must guarantee that the expected utility that the agent receives is at least as much as their threshold wage. The agent's utility is the utility it receives from the incentive scheme less their disutility of exerting effort a. Let us make this precise:

$$x = \text{ agent's output}$$

$$a = \text{ the agent's effort level}$$

$$I(x) = \text{ the payment schedule to the agent}$$

$$f(x, a) = \text{ the p.d.f. of output } x \text{ and effort level } a$$

$$G(s) = \text{ principal's utility for income } s$$
$$U(s) = \text{ agent's utility for income } s$$
$$V(a) = \text{ agent's disutility for choosing effort level } a$$
$$E(\cdot) = \text{ expectation (of utility)}$$
$$K(\cdot) = \text{ agent's reservation utility,}$$

which is the utility the agent can get by working elsewhere.

Based on the above assumptions, the principal's problem becomes[2]

$$\max_{I(x),a} E[G(x - I(x)] = \max_{I(x),a} \int G\left(x - I(x)\right) f(x,a)dx, \text{ subject to} \tag{2a}$$

$$E[U(I(x)) - V(a)] = \int U\left(I(x) - V(a)\right) f(x,a)dx \geq K, \text{ and} \tag{2b}$$

$$\forall a' \neq a, E\left[U(I(x)) - V(a)\right] \geq E\left[U(I(x)) - V(a')\right]. \tag{2c}$$

This says that

1. The principal wants to choose $I(x)$ to maximize their expected utility, subject to the constraint that
2. the agent's utility must be at least as much as its reservation utility and to the constraint that
3. effort level a is more desirable (i.e., results in at least as much payment) to the agent than any other effort level.

The choice variables in the model are $I(x)$ and a. The principal chooses $I(x)$ to maximize its utility under the given constraints and the agent chooses a to maximize its own utility. The other parameters are exogenous in that they are not determined by the principal or agent.

3 Applying principal agent theory to cloud computing

As in any business model, when a cloud consumer agrees to pay for the services of a cloud provider, both parties enter a contractual relationship. However, typical cloud computing contracts are heavily one-sided that describe generic cloud provider responsibilities in terms of performance, service levels, etc. and detailed requirements from consumers with respect to security, payments, and associated penalties. We expect that different approaches such as the principal agent theory model presented in the previous section can be used to make these contracts more customer-specific and provide customization in terms of additional security that satisfies the requirements of the consumer. In this section, we present the mathematical requirements of our paradigm for relating principal agent theory to cloud computing. We briefly describe each area, and provide the necessary skeleton to work from.

As mentioned in section 2, we view the relationship between a cloud provider and cloud consumer as a principal agent relationship. We define the principal agent relationship in cloud computing as follows: The cloud consumer utilizes a third party (i.e., cloud provider) to provide various cloud services such as infrastructure and/or services and host consumer applications and data. The cloud provider is acting as an agent for the cloud consumer, the principal. While not having to manage lower-level infrastructure components releases the management burden for and creates cost savings for a consumer, it now has to trust the provider to manage these components in a manner that guarantees the security of the consumer's data and applications. Security is not a simple concept and cannot be easily measured. The three main security objectives of confidentiality, integrity, and availability have to be preserved. Our objective is to use

[2]We do not get into the details of the domain of integration in this paper. All integrals are assumed to be definite integrals taken over the largest domain possible. The details can be found in [17].

principal agent theory to examine ways to motivate a possibly risk-averse cloud provider to accept certain risks associated with cloud management tasks and once the provider agrees to accept the task (i.e., contract is signed), to be able to trust that the provider will exert the level of effort desired by the principal. Using the model formulated in the previous section, we discuss the various parameters that have to be considered in applying a principal agent paradigm to cloud computing.

3.1 The Agent's Effort

The agent's level of effort is a. In cloud computing, the cloud consumer desires that the cloud provider take every precaution in managing the infrastructure and security components. These precautions include having a secure facility, training and educating personnel, monitoring its own employees, prompt updating of patches and vulnerabilities, and keeping backups and audit records. In other words, the agent's effort level can be the monetary and/or physical effort that the cloud provider puts in to complete the task or provide the service. This level of effort will differ from agent to agent and will also be dependent on the cloud deployment model used. Comparing various effort levels of a cloud provider and among many providers is facilitated if the level of effort can be converted to a monetary value.

3.2 The Agent's Output

The agent's output is x. When the cloud provider exerts an effort to maintain its infrastructure, that effort results in some sort of output that produces mutual beneficial payoffs to both the cloud provider and consumer. Since the cloud consumer cannot directly monitor all of the provider's efforts, the output should be something that, while can be easily monitored, is closely tied to the agent's effort. This output, or its result, is shared between the provider and consumer through the incentive scheme. In cloud computing, the output is not an easily monitored value. The ideal outcome (for the cloud consumer) is that for the duration of the contract, the consumer's data and applications stored on the cloud are secure. That is, the three properties of confidentiality, integrity, and availability are maintained. How to detect, measure and verify this is a difficult problem. Furthermore, the security of the data itself is not proportional to a possible security threat. A relatively small risk can result in a catastrophic failure, while a large-scaled attack may result in only minimal damage to the infrastructure and data. While the output is difficult to verify, assuming it can be done, one possible way to measure the output is to calculate the value of the consumer's data. If data has been compromised in any way, then the data is no longer valuable to its owner. If the value of the data was not very important to begin with, losing the data will have no severe consequence to the consumer, hence she will not have to offer a steep incentive scheme. Also, value can be a monetary measure, providing a means to compare parameters. The value of a consumer's data can in theory be unbounded. On the other hand, when data value is used for x, incentive schemes may not make much sense. After all, the consumer is willing to pay the provider to use its resources for the convenience and potential savings it would have from not having to invest in and maintain its own infrastructure. A more accurate model should reflect this.

3.3 The Probability Density Function

The probability density function is $f(x, a)$. We have many modeling choices for $f(x, a)$, however, they must be tied to reality. In [17], he discusses the exponential distribution, and also shows how the problem can be discretized and a probability mass function used instead of a pdf. For the exponential distribution [17] uses a as the parameter in the exponential distribution; that is, $f(x, 1/a) = ae^{-ax}$ for non-negative x, see Fig. 2. Of course, the exponential distribution is useful when modeling outputs that have the memoryless property. If using an exponential distribution, we would need to determine if the cloud provider's output is memoryless. We are considering generalizing to the gamma distribution, when they may be other physical parameters to consider. The gamma distribution has been used to model the time required to service customers in queuing systems, the lifetime of devices and systems in reliability studies, and the defect of clustering behavior in VLSI chips [23].

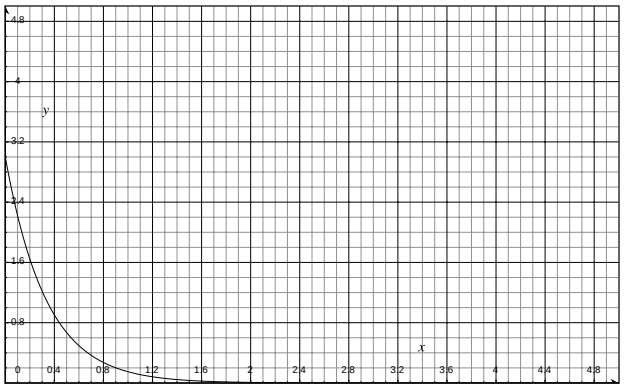

Figure 2: $f(x, 1/3) = 3e^{-3x}$

3.4 The Incentive Scheme

$I(x)$ is the incentive scheme offered to the agent. In principal agent theory, most compensation schemes are affine functions of the output x. For example, the general affine form is $I(x) = A + Bx$, where A is a fixed payment that guarantees the agent will receive some payment even when circumstances not under the agent's control results in worst case results (e.g., the region that the cloud is hosted in suffers a nuclear attack and everything is demolished), see Fig. 3. This fixed component is required because we assume that the agent is risk-averse. On the other hand, having just a fixed payment creates disincentive for the provider to work hard, forcing the cloud consumer to bear all the associated risk. The output-proportional component allows the consumer and provider to share risk, creating a better trust relationship.

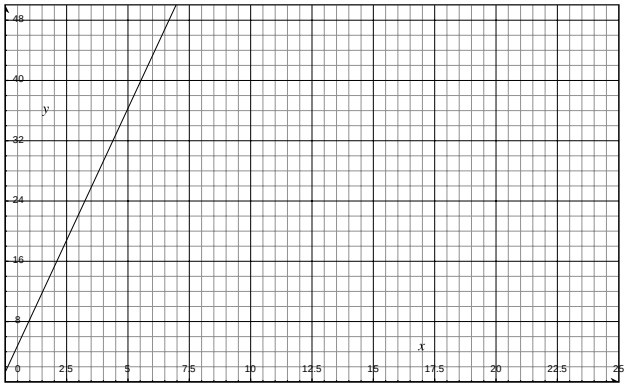

Figure 3: $I(x) = 1.3 + 7x$

Current pricing schemes in the cloud computing industry are 'pay-as-you-go' in which you only pay for the time and amount of resources used instead of paying a flat fee. This can be beneficial to cloud consumers, particularly those that have bursts of activity. Since this payment structure affects the reservation utility as well, we need to create an incentive scheme that provides some level of pay-as-you-go, but still has a component that is tied to the agent's output. In fact, the incentive scheme can be a component of pay-

as-you-go plus an output-proportional component. The new pay-as-you-go component would have to be considerably less than the current one in order to maximize the consumer's utility, while motivating the provider to exert a desired level of effort. If we just denote the cloud provider's current pay-as-you-go scheme as Y. We can simply state the incentive scheme as $I(x) = Y\beta + Bx$, where $0 < \beta < 1$ is a proportionality factor. This is not as simple as it seems. Current payment schemes charge on a monthly basis, while incentive contracts generally last longer (i.e., yearly). These different nuances need to be taken into account when considering what form the incentive scheme should take.

3.5 Utility functions

Utility functions are variables that reflect an individual's attitude toward risk. They assign numerical values to potential outputs in a way that the values rank the outputs according to the decision-maker's preference. Risk-neutral individuals have a straight line utility function, risk-averse individuals have concave utility functions and risk-prone individuals have convex functions. To model our risk-averse entities, we could use a simple utility function such as $U(x) = \log_b(x)$. However, since utilities are different for each entity, and even for different situations for one entity, will suggest using a general utility function that has a risk aversion parameter. Substituting different values for this parameter will give us the different utility functions we need to analyze the relationship between risk attitudes and incentive schemes. One option to consider is the exponential utility function (see Fig. 4):

$$U(x) = 1 - e^{-xR}$$

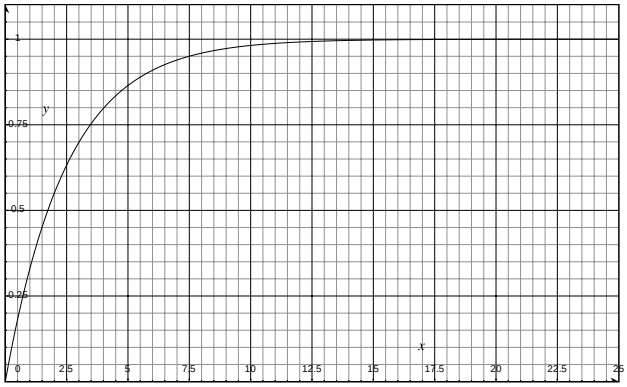

Figure 4: $U(x) = 1 - e^{-x\cdot(.4)}$

This assumes that if risk-averse, an individual has a constant absolute risk aversion utility function. Thus the utility functions will take the form of $G(s) = 1 - e^{-sR_1}$ for the principal's utility function, and $U(s) = 1 - e^{-sR_2}$ for the agent's utility function, where R_1 and R_2 are the principal's degree of risk aversion and the agent's degree of risk aversion, respectively. In addition, $0 < R_i < 1$ and the party is more risk-averse as $R_i \to 0$. The agent's *disutility* of effort $V(a)$ is a function of effort and can generally be represented as follows [3] (see Fig. 5).

$$V(a) = a^2.$$

By giving different values for R_1 and R_2, we can account for the different risk attitudes of the cloud consumer and provider.

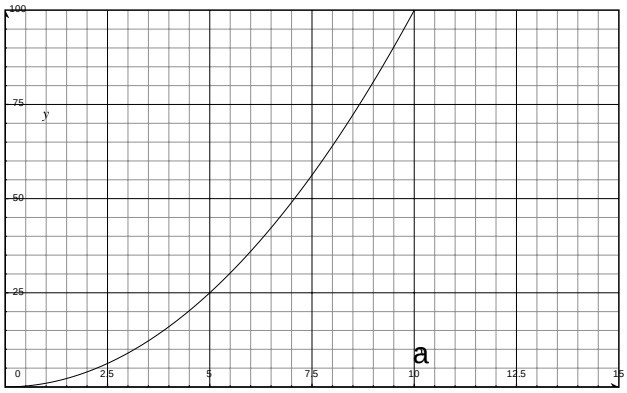

Figure 5: Heuristic: $V(a) = a^2$

Table 1: Heuristic Payoff Matrix

	Cloud provider exerts effort a_1	Cloud provider exerts effort a_2
Cloud consumer offers incentive scheme I_1	$x_1 - I_1$, I_1	$x_2 - I_1$, I_1
Cloud consumer offers incentive scheme I_2	$x_1 - I_2$, I_2	$x_2 - I_2$, I_2

3.6 Reservation Utility

The reservation utility K is the utility that the agent can get by working elsewhere. The reservation utility of an agent can vary immensely from agent to agent, particularly in a pay-as-you-go utility computing model such as cloud computing. In fact, the amount a principal would pay in a pay-as-you-go model would also differ from principal to principal for the same cloud provider.

3.7 Brief Discussion

Summarizing our discussion up to this point, we recognize that many coefficients have to be specified: the utility functions (two parameters: R_1 and R_2), the probability distribution (two parameters: x and a), the agent's level of effort, and the agent's reservation utility K. It is seemingly impossible to substitute specific values for these coefficients and obtain a tangible value for general cases of cloud computing. This is because the nature of the model relies heavily on statistical data and user preference, which varies from agent to agent and principal to principal [3]. Also, incentive schemes are currently not used in cloud computing system so there is no previous data to compare with.

3.8 Game Theory

Principal agent theory has also been defined as a class of games of *imperfect information* in which one player (the principal) attempts to offer *incentives* to the other (the agent) to encourage the agent to act in the principal's best interest [30]. The game aspect of principal agent theory focuses on the moral hazard problem as well. In terms of game theory, principal agent theory involves changing the rules of the game so that the self-motivated rational choices of the agent align with what the principal desires.

If we think of the principal agent problem as a game, then our objective is to reach a Nash equilibrium where each player is assumed to know the equilibrium strategies of the other players, and no player has anything to gain by changing only his or her own strategy unilaterally [29]. Thus, if each player chooses a strategy and neither player can benefit by changing its strategy while the other player keeps their strategy unchanged, then the current set of strategy choices and the corresponding payoffs constitute a Nash equilibrium. In our case, we would want to devise a contract in such a way that if the cloud consumer offered a contract I_i and cloud provider exerted an effort level of a_j, then a_j would result in the level of effort that would provide the most utility to both parties under that contract and hence would be a desirable level of effort for the consumer as well as reach Nash equilibrium. A trivial payoff matrix is shown in Table 1. The rows of the matrix represent the moves the cloud consumer can make while the columns represent the cloud producer's moves. In each box, the first number is the payoff received by the consumer; the second is the payoff for the provider. For example, when the consumer offers incentive scheme I_i, and the provider chooses an effort level of a_j, this results in an output of x_j. This output is shared between the cloud consumer and provider by giving a payoff of $x_j - I_i$ to the consumer and a payoff of I_i to the provider.

However, a Nash equilibrium does not necessarily mean the best cumulative payoff for all the players involved. Therefore, a Nash equilibrium analysis of principal agent theory would be useful if the consumer had a priori knowledge that the provider would exert effort level a_j and would offer a contract with an incentive scheme that gave the principal the best payoff if a_j is exerted by the provider. While current payment structures are pay-as-you-go, using information from white papers (Amazon EC2/S3 security services, [1]) and historical/empirical evidence of management of systems, we may be able to estimate the level of effort given for each company, and counter with an offer that encompasses a pay-as-you-go scheme with a Nash equilibrium. We realize this is controversial and especially request feedback at this point.

4 Open Discussion

The underlying assumptions in principal agent theory are much more complicated than presented here. Various different models with varying degrees of mathematical complexity arise and have been studied, depending on the various assumptions made about the principal and the agent. In this paper, we have proposed and used a relatively standard model to present/suggest this particular theory as a new paradigm for achieving trust and balancing risk between a cloud consumer and provider. It is our argument that current techniques and technologies by themselves will be insufficient to solve this problem. The user of incentive-based contracts can help to balance the power structure and information asymmetry in the consumer-provider relationship. Until this can be done (by whatever means), cloud computing will not be a successful long term viable alternative.

As has been hinted at, capturing a principal agent relationship is not that simple. The cloud provider, while acting as an agent to the consumer, has another principal agent relationship where it is the principal and its employees are its agents. Depending on how the cloud is used, there may be a cloud infrastructure provider who owns and manages the underlying resources, a cloud provider who launches virtual machines and develops applications on the cloud, and the cloud consumer who uses these applications to perform its own tasks (i.e., a three-tiered relationship). The relationship between the principal and agent does not always involve just the two parties. For example, in cloud computing, consumers can utilize different providers to host services and data. Cloud providers themselves are multi-tenant, providing services to more than one principal. Principal agent theory recognizes this and the literature covers single-principal single-agent, single-principal multiple-agent, multiple-principal single-agent, and multiple-principal-multiple-agent models each with different incentive scheme structures depending on the nature of the task. For example, the single-principal multiple-agent model may take into consideration the fact that agents may collude among themselves. In such cases, agent performance and payoff schemes many be measured by ranking the outputs or paying only the agent with the best effort. Also Levitt [24] states that to induce higher levels of effort from the agent the principal must provide stronger incentives for each agent, but when only the best agent's output is rewarded, the benefit to the principal is marginal. Therefore, as the number of agent's increase, the incentive schemes will be powered lower, reducing the effectiveness of the incentives. If a cloud consumer

were to use multiple providers to host its data and applications, these issues would need to be taken into consideration. Also, the longer the consumer and producer maintain a relationship (i.e., contract is renewed), the more information the consumer possesses about the cloud provider. Therefore, long-term relationships can result in incentive contracts that grow less steep over time.

Another factor that may affect the incentive mechanism is the type of cloud deployment model used. For example, in a PaaS deployment model the provider has control over the majority of the infrastructure. This implies that the provider has more knowledge of the level of effort required to maintain the system and that the level of effort is even more difficult to monitor compared to an IaaS-type deployment model. Therefore, the incentive schemes for each deployment model may require different approaches.

Early on, we stated that when providers' security mechanisms fail, the consumer's data is at risk. While this is true, attacks against the infrastructure can also damage the provider. Therefore, the provider has some incentive not introduced by the principal, to manage its infrastructures securely.

5 Conclusions

In this paper we have attempted to provide a new paradigm for trust and risk in computer security issues. We presented principal agent theory as a way to accomplish this. While we use cloud computing as a specific example, we believe this paradigm can be applied to any computer security issue in which a principal agent relationship exists. On the other hand, we do not want to leave the reader with the opinion that principal agent theory is a panacea for all relationships where asymmetric information exists. To be intellectually honest, we note that it has been criticized by [14, 28] as having the following limitations.

5.1 Limitations of Our New Paradigm

Modeling the real world requires assumptions and simplifications that may not provide a true representation of real world relationships. Some estimation is also required to determine parameters of the model, which may result in inaccurate reflections of the nature of the principal agent relationship. In the example we provided, the risk attitudes of the parties and the agent's threshold value were examples of parameters that had to be estimated. Principal agent theory itself also possesses opponents. It treats individuals as extremely rational utility-maximizing beings and focuses solely on monetary incentives. In cloud computing, the agent and principal are likely not humans but corporations, so that this assumption may work well. However, even companies act from complex motivations ranging from altruism, responsibility, and the need for recognition. Another limitation that has been mentioned is that principal agent theory only looks at the obligation of the agent to the principal, and ignores the principal's obligations to the agent. In cloud computing, while the cloud consumer expects the provider to manage the infrastructure properly, the consumer uses its own local devices to connect to cloud resources. These local devices, if not properly secured, can cause malicious software to enter the cloud, causing angst for both the provider and the other tenants within the cloud. Therefore, the consumer, as a principal has an obligation to the provider, its agent, to manage its own local devices securely as well.

5.2 Last Word

Realizing that the current technical and procedural approaches to balancing risk and security are not sufficient in themselves, we proposed principal agent theory as a new paradigm to balance trust and risk. Cloud computing has been called a multi-tenant problem, comparing the issues of trust arising in 'leasing' cloud provider's infrastructures and the fact that many consumers 'share' these resources together to a problem of a landlord renting land or housing to multiple tenants [7]. Principal agent theory is also known as the landlord-tenant relationship due to its popular application in tenancy [32]. Despite the limitations discussed above, the theory can be useful in identifying factors that lead to a lack of trust and provide ways to model different ways of decreasing the effects of these factors, thereby increasing trust in tenancy-related computer security problems.

In conclusion, our new paradigm can be summed up as — money talks. Cloud providers are not charitable organizations willing to host others' data and applications from the sheer goodness of their hearts. The only way we can be assured that they are working towards the consumers' goals is to "improve their bottom line." Lastly, it is important to stress that incentives by themselves are not sufficient in securing one's data in the clouds. Incentives help balance trust and risk. It spreads the risk between the cloud provider and consumer by incentivizing the provider to accept some risk. Knowing that the provider is willing to take some risk, in the form of monetary payments, and hence responsibility, allows the consumer to better trust the cloud provider. However, regardless of one's best efforts, things can and do go wrong. Therefore, consumers would still need to employ various suitable security measures such as data encryption, data backup, virtual machine hardening, and monitoring.

6 Acknowledgements

We are grateful to Shulamit bat Yehudit for her helpful comments.

References

[1] Amazon.com. Amazon web services: Overview of security processes. Technical report, September 2008.

[2] Amazon.com. Extend your infrastructure with amazon VPC, December 30 2009.

[3] R. A. Androkovich. *The impact of risk on contract structure in a principal-agent model : an application to the Alberta sugar beet industry*. PhD thesis, 1986.

[4] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. H. Katz, A. Konwinski, G. Lee, D. A. Patterson, A. Rabkin, I. Stoica, and M. Zaharia. Above the clouds: A Berkeley view of cloud computing. Technical report, UC Berkeley Reliable Adaptive Distributed Systems Laboratory, February 10 2009.

[5] K. J. Arrow. Insurance, risk and resource allocation. *Essays in the Theory of Risk-Bearing*, pages 134–143, 1971.

[6] R. Axelrod. *The Evolution of Cooperation*. Basic Books, New York, NY, 1984.

[7] D. Banks, J. Erickson, and M. Rhodes. Multi-tenancy in cloud-based collaboration services. Technical report, HP Labs, February 21 2009.

[8] D. Besanko, D. Dranove, M. Shanley, and S. Schaefer. *Economics of Strategy*. 3rd edition, 2003.

[9] B. Blakley. The emperor's old armor. In *NSPW'96: Proceedings of the 1996 New Security Paradigms Workshop*, pages 2–16, New York, NY, USA, 1996. ACM.

[10] J. Brodkin. Loss of customer data spurs closure of online storage service. *Network World*, August 11 2008.

[11] J. Brodkin. Survey casts doubt on cloud adoption. *Network World*, June 26 2009.

[12] C. Cachin, I. Keidar, and A. Shraer. Trusting the cloud. *SIGACT News*, 40(2):81–86, 2009.

[13] Cloud Security Alliance. Security guidance for critical areas of focus in cloud computing, ver. 2.1. Technical report, 2009.

[14] K. M. Eisenhardt. Agency theory: An assessment and review. *The Academy of Management Review*, 14(1):57–74, 1989.

[15] D. H. Guston. Principal-agent theory and the structure of science policy. *Science and Public Policy*, 23(4):229–240, 1996.

[16] J. Heiser and M. Nicolett. Assessing the security risks of cloud computing. Technical report, Gartner, 2008.

[17] B. Holmström. Moral hazard and observability. *The Bell Journal of Economics*, 10(1):74–91, 1979.

[18] M. Jensen, J. Schwenk, N. Gruschka, and L. L. Iacono. On technical security issues in cloud computing, 2009.

[19] A. Joch. Cloud computing: Is it secure enough? *Federal Computer Week*, June 18 2009.

[20] A. Jøsang. The right type of trust for distributed systems. In *NSPW '96: Proceedings of the 1996 New Security Paradigms Workshop*, pages 119–131, New York, NY, USA, 1996. ACM.

[21] J.-J. Laffont and D. Martimort. *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press, Princeton, NY, illustrated edition, 2001.

[22] R. G. Lennon, L. A. Skår, M. Udnæs, A. J. Berre, A. Zeid, D. Roman, E. Landre, and W.-J. van den Heuvel. Best practices in cloud computing: designing for the cloud, 2009.

[23] A. Leon-Garcia. *Probability and Random Processes for Electrical Engineering*. Addion-Wesley, Reading, Massachusetts, 1994.

[24] S. D. Levitt. Optimal incentive schemes when only the agents' "best" output matters to the principal. *The Rand Journal Of Economics*, 26(4):744–760, 1995.

[25] P. Mell and T. Grance. Effectively and securely using the cloud computing paradigm, October 7 2009.

[26] P. Mell and T. Grance. The NIST definition of cloud computing (ver. 15). Technical report, National Institute of Standards and Technology, Information Technology Laboratory, October 7 2009.

[27] C. Metz. DDoS attack rains down on amazon cloud. *The Register*, October 5th 2009.

[28] T. M. Moe. The new economics of organization. *American Journal of Political Science*, 28(4):739–777, 1984.

[29] J. F. Nash. Equilibrium points in N-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49, 1950.

[30] E. Rasmusen. Moral hazard: hidden actions. In E. Rasmusen, editor, *Games and information: an introduction to game theory*, page 560. Wiley-Blackwell, Malden, MA, 2007.

[31] S. A. Ross. The economic theory of agency: The principal's problem. *The American Economic Review*, 63(2):134–139, 1973.

[32] A. Smith. *The Wealth of Nations*. The Modern Library, New York, 1776.

[33] L. M. Vaquero, L. Rodero-Merino, J. Caceres, and M. Lindner. A break in the clouds: Towards a cloud definition. *ACM SIGCOMM Computer Communication Review*, 39(1):50–55, 2009.

[34] Wikipedia. Cloud computing. `http://en.wikipedia.org/wiki/Cloud_computing`.